RESEARCH AND DEVELOPMENT TECHNICAL REPORT
CECOM-TR-94-12

THE USE OF TIME SERIES FORECASTING IN CONTRACTOR PERFORMANCE ANALYSIS

RICHARD J. D'ACCARDI, Ph.D.
COST ANALYSIS DIVISION
DIRECTORATE OF RESOURCE MANAGEMENT

November 1994

DISTRIBUTION STATEMENT

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION IS UNLIMITED.

CECOM
U.S. ARMY COMMUNICATIONS-ELECTRONICS COMMAND
DIRECTORATE OF RESOURCE MANAGEMENT
FORT MONMOUTH, NEW JERSEY 07703-5034

NOTICES

Disclaimers

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
November 1994

3. REPORT TYPE AND DATES COVERED
Technical Report - 1994

4. TITLE AND SUBTITLE
THE USE OF TIME SERIES FORECASTING IN CONTRACTOR PERFORMANCE ANALYSIS

5. FUNDING NUMBERS

6. AUTHOR(S)
Richard J. D'Accardi, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Communications-Electronics Command (CECOM)
Directorate of Resource Management
Cost Analysis Division, ATTN: AMSEL-CP-CA-CS
Fort Monmouth, NJ 07703-5034

8. PERFORMING ORGANIZATION REPORT NUMBER
CECOM-TR-94-12

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

19941205 098

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
The application of time series forecasting is presented as a method of estimating either contract Estimates-at-Completion (EAC) or year-ahead contract costs based on information contained in monthly Contract Performance Reports (CPR). Considering the premise that the pattern of a contractor's expenditures becomes characteristic as the contract proceeds, a procedural approach is recommended for characterizing the cumulative Actual Cost of Work Performed (ACWP) using time series modeling. The recommended technique stresses the structuring of appropriate difference equations through the criterion of minimum residual variance. The main objectives of this report are: (1) Development of a procedure for structuring forecasting models from non-stationary stochastic realizations. This specifically covers the identification and fitting of Autoregressive (AR), and integrated Autoregressive-Moving Averages (ARIMA) models. (2) Illustration of the validity of the appropriate model through a sensitivity analysis using CPR information from two Communications-Electronics Command contracts.

14. SUBJECT TERMS
Time series; time series modeling; time series forecasting; cumulative ACWP; ARIMA; Autoregressive; minimum residual variance; Estimates-at-Completion; non-stationary time series

15. NUMBER OF PAGES
22

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

CONTENTS

FIGURES

FIGURES (Continued)

TABLES

THE USE OF TIME SERIES FORECASTING IN CONTRACTOR PERFORMANCE ANALYSIS

1.0 INTRODUCTION

This report deals with the application of time series forecasting as a method of estimating either contract Estimates-at-Completion (EAC) or year-ahead contract costs based on information contained in monthly Contract Performance Reports (CPR). Specifically, the main thrust is to obtain accurate short-term forecasts of cumulative Actual Cost of Work Performed (ACWP). The method presented herein can also be applied to the cumulative Cost Variance (CV). The latter takes a more direct look at an expected cost overrun.

Due to the nature of the cumulative ACWP and the time dependence of contract expenditures, a logical approach for more accurately forecasting this function seems to lie within the realm of time series modeling. Considering the premise that the pattern of a contractor's expenditures becomes characteristic as the contract proceeds, a procedural approach is recommended for characterizing the cumulative ACWP utilizing time series modeling. The recommended technique stresses the structuring of appropriate difference equations through the criterion of minimum residual variance. The main objectives of this report are as follows:

(1) Development of a procedure for structuring forecasting models from nonstationary stochastic realizations. This specifically covers the identification and fitting of Autoregressive (AR) and integrated Autoregressive-Moving Averages (ARIMA) models.

(2) Illustration of the validity of the appropriate model through a sensitivity analysis utilizing CPR information from two Communications-Electronics Command contracts.

To achieve these objectives, Section 2.0 of this report presents a brief view of the concepts of time series and provides a procedural approach for obtaining time series models. Section 3.0 contains the identification of the CPR data which includes a determination of whether or not the data are nonstationary, the fitting of an appropriate model to the data, and the checking of the fit of the models. Section 4.0 contains conclusions and a comparison of time series forecasts to those of Performance Analyser.

2.0 CONCEPTS IN TIME SERIES

Any phenomenon that changes with time, and any collection of data measuring some aspect of such a phenomenon, can be considered a time series. Time series can either be deterministic or nondeterministic functions of an independent variable, usually time. In most instances, however, they will be nondeterministic functions. A nondeterministic function exhibits random or fluctuating properties, and, hence, it is not possible to exactly forecast its future values; in other words, such time series can be described only by statistical laws or models. We assume that we may describe a time series at a given time *t* by a random variable and its associated probability distribution. Thus we may describe the behavior of a time series at all instances by an ordered set of random variables and the associated probability distributions, denoted by $\{X_t\}$ and $\{f_{xt}\}$, t = 0, ±1, ±2,...... Such an ordered set of random variables is called a stochastic process. Thus, an observed time series, x_t, can be considered as <u>one</u> realization of an infinite ensemble of functions which may have been generated by a stochastic process. A stochastic process is said to be <u>strictly</u> stationary if the joint distribution of any set of observations is unaffected by shifting all times of the observations ahead or backward by any integer amount *k* [1]. A stationary stochastic process may be described in terms of its mean μ which is estimated by:

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t \, , \tag{2.1}$$

its variance σ^2 which is estimated by:

$$s_x^2 = \frac{1}{n}\sum_{t=1}^{n}\left(x_t - \bar{x}\right)^2 , \tag{2.2}$$

its sample autocovariance function, which measures the extent to which two random variables are linearly independent:

$$c_{xx}(k) = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x}), \quad \text{k = 0, 1, ..., n-1} \tag{2.3}$$

and the sample autocorrelation function, which acts like a correlation coefficient:

$$r_{xx}(k) = c_{xx}(k) / c_{xx}(0), \qquad \text{k = 0, 1, ..., n-1} \tag{2.4}$$

2.1 Stationary and Nonstationary Time Series

A stationary time series is one which is in statistical equilibrium in the sense that its properties do not change with respect to time, whereas a nonstationary time series is one whose properties change with time. Time series occurring in practice are usually nonstationary in nature and can be divided into three classes:

(1) Those that exhibit stationary properties over a long period of time.

(2) Those that are approximately stationary over very short periods of time.

(3) Those that exhibit nonstationary properties, that is, their visual properties change continuously with time.

At present, there exist techniques to analyze stationary time series, but the techniques available for the analysis of nonstationary time series are inadequate and do not lend themselves to meaningful interpretations of physical problems. However, nonstationary time series can be *adjusted* so that the existing techniques of stationary time series analysis can be applied. Adjustments are accomplished by applying a proper filter to the observed nonstationary time series to remove the nonstationary components.

The selection of a proper filter is accomplished through a search for a mathematical function that will transform a nonstationary time series into a stationary time series. One of the most often used and most efficient methods of filtering is through the application of difference equations [1, 2]. A first-order difference equation is defined by:

$$y_t = x_t - x_{t-1}, \tag{2.5}$$

where x_t is the observed nonstationary time series and y_t is the first-difference series. Similarly, a second-order difference equation is defined by:

$$w_t = x_t - 2x_{t-1} + x_{t-2}, \tag{2.6}$$

and so on. A first- or second-order difference equation will usually be sufficient to transform most practically occurring nonstationary time series [1].

To identify whether or not the observed series exhibits stationary or nonstationary properties, one can use certain data analysis tools. In addition to graphical representations of the observed series, the sample autocorrelation function of the observed series and a trend test applied to the observed series are important.

For the observed series and its first and second differences, the sample autocorrelation functions (equation 2.4) are computed and a trend test, i.e., Kendall's Tau [3], is performed. (*For a stationary time series, the sample autocorrelation function has the property that it dampens out fairly rapidly and it contains no trend.*) Following these procedures, sufficient information can be obtained to determine if the observed series exhibits either stationary or nonstationary properties, and whether or not a first- or second-order difference equation will remove the nonstationarities.

Once a model for the stationary series is obtained, a *backward filter* is applied to the *fitted stationary model* so that future values of the observed series can be forecasted.

2.2 Parametric Time-Series Models

To be able to forecast values for an observed series, we fit parametric time series models, either autoregressive, moving-average, or a combination of the two. These stationary stochastic models assume the process (series) remains in equilibrium about a constant mean level. The general autoregressive process is given by:

$$x_t - \mu = \alpha_1(x_{t-1} - \mu) + \cdots + \alpha_m(x_{t-m} - \mu) + Z_t \tag{2.7}$$

where μ is the mean of x_t, Z_t is a purely random process [2], and m is the order of the process. The general moving-average process is given by:

$$x_t - \mu = Z_t - \beta_1 Z_{t-1} - \cdots - \beta Z_{t-q}, \tag{2.8}$$

where μ and Z are as defined in equation 2.6, and q is the order of the process. The general mixed autoregressive-moving averages process is given by:

$$x_t - \mu = \alpha_1(x_{t-1} - \mu) + \cdots + \alpha_m(x_{t-m} - \mu) + Z_t - \beta_1 Z_{t-1} - \cdots - \beta_q Z_{t-q}, \tag{2.9}$$

where q is independent of m.

We shall now consider the criterion for selecting the process, its order (which gives the best fit to an observed series), the procedure to estimate its parameters, a diagnostic check of goodness-of-fit, and how the model can be used in forecasting.

2.2.1 Selecting the Best Model

The criterion for selecting the order of the process (that which will give the best fit) is the residual variance for the different orders of the parametric models fitted to the data.

The *residual variances* are computed and plotted against the order of the process; the minimum residual variance will correspond to the correct order for the process. After this has been done for the autoregressive, moving-average, and the mixed autoregressive-moving average processes, we compare the minimum residual variances. The *minimal* one will correspond to the process (and its order) that will give the best fit to the data. When one fits a model to a given set of observations, the principle of *parsimony* should always be considered. That is, the least number of parameters should be employed to obtain an adequate representation [1].

2.2.2 Estimation of Parameters

To obtain the residual variances above, we first estimate the parameters for the different orders of each individual process. To estimate the parameters for the autoregressive process [2], we first assume that the Z_t process is *normal* with zero mean and variance σ_z^2. Then the log-likelihood function for fixed m, conditional on the values $x_1, x_2, \cdots, x_m$, can be expressed as:

$$\ell(\mu,\alpha_1,\cdots,\alpha_{m\beta} | x_1,\cdots,x_m) = \left(n-m)\left(\ln\sqrt{2\pi+\ln\sigma_z}\right) - \frac{1}{2\sigma_z^2}\sum_{t=m+1}^{n}\left[(x_t-\mu)-\alpha_1(x_{t-1}-\mu)-\cdots-\alpha_m(x_{t-m}-\mu)\right]^2 \right) \tag{2.10}$$

For estimating the parameters $\mu, \alpha_1, \cdots, \alpha_m$, we need only consider the sum of squares function:

$$S(\mu,\alpha_1,\cdots,\alpha_m | x_1,\cdots,x_m) = \sum_{t=m+1}^{n}\left[(x_t-\mu)-\alpha_1(x_{t-1}-\mu)-\cdots-\alpha_m(x_{t-m}-\mu)\right]^2 \tag{2.11}$$

now assuming that $\hat{\mu}$ may be approximated by $\bar{x}$ and that the sample autocorrelation function (equation 2.3) can be written as:

$$c_{xx}(j) \cong \sum_{t=m+1}^{n}(x_t-\hat{\mu})(x_{t-j}-\hat{\mu}) \qquad j=1,\cdots,m, \tag{2.12}$$

then the maximum likelihood equations may be expressed as:

$$c_{xx}(j) = \hat{\alpha}_1 c_{xx}(j-1) + \hat{\alpha}_2 c_{xx}(j-2) + \cdots + \hat{\alpha}_m c_{xx}(j-m), \tag{2.13}$$

where $j = 1,2,\cdots,m$. Solving the m simultaneous equations, we obtain the estimates $\hat{\alpha}_1,\cdots,\hat{\alpha}_m$. The residual sum of squares may be expressed as:

$$S(\hat{\mu},\hat{\alpha}_1,\cdots,\hat{\alpha}_m) \cong (n-m)\left[c_{xx}(0)-\hat{\alpha}_1 c_{xx}(1)-\cdots-\hat{\alpha}_m c_{xx}(m)\right] \tag{2.14}$$

and the residual variance by:

$$s_z^2 = \frac{1}{n-2m-1} S(\hat{\mu}, \hat{\alpha}_1, \cdots, \hat{\alpha}_m). \tag{2.15}$$

To estimate the parameters of the moving average process and the mixed autoregressive-moving average process, we use a numerical technique to recursively build the log-likelihood function [2]. By varying the values of the parameters (usually between -1 and +1), we can search for the parameter estimates which minimize the sum of squares function for each process. For example, for the general moving average process the sum of squares function is given by:

$$S(\hat{\mu}, \hat{\beta}_1, \cdots, \hat{\beta}_q) = \sum_{t-q}^{n} z_t^2, \tag{2.16}$$

where

$$z_t = x_t - \mu + \beta_1 z_{t-1} + \cdots + \beta_q z_{t-q},$$

and $z_t = 0$ for $t < q$. The residual sum of squares is given by:

$$S_z^2(q) = S(\hat{\mu}, \hat{\beta}_1, \cdots, \hat{\beta}_q) / (n-q-1). \tag{2.17}$$

Similarly, the residual sum of squares for the mixed autoregressive-moving average process can be expressed as:

$$s_z^2(m,q) = \frac{1}{n-2m-q-1} S(\hat{\mu}, \hat{\alpha}_1, \cdots, \hat{\alpha}_m, \hat{\beta}_1, \cdots, \hat{\beta}_q). \tag{2.18}$$

2.2.3 Checking the Fit

Once a model is fitted to the stationary series, the adequacy of the model must be determined. If it was necessary to *filter* the observed series, the first step is to apply a *backward filter* of the same form so that the fitted model represents the observed series. Thus, with the *backward filter* inserted, the fitted model will simulate the behavior of the observed series. Then the residuals, the observed series minus the modeled series, should behave approximately like a purely random process (white noise). That is, the sample autocorrelation function (equation 2.3) should effectively be zero for all lags except the zeroth. To determine the fit, a test for white noise is applied [2].

2.2.4 Forecasting

After checking the fit, the resulting equation can be used to forecast a value $x_{t+\ell}$, $\ell \geq 1$, when we are currently at time t. This forecast is said to be made at origin t for a lead time ℓ. The minimum mean square error forecast [1] for any lead time ℓ is given by the conditional expectation $E_t x_{t+\ell}$, of $x_{t+\ell}$ at origin t, given knowledge of all the x's up to time t. That is:

$$\hat{x}_t(\ell) = E_t[x_{t+\ell}]. \tag{2.19}$$

The required conditional expectation occurring in the forecasting models can be found using Box and Jenkins [1]:

$$E_t[x_{t+j}] = \hat{x}_t(j), \qquad E_t[z_{t+j}] = 0, \qquad j = 1, 2, \tag{2.20}$$

and

$$E_t[x_{t-j}] = x_{t-j}, \qquad E_t[z_{t-j}] = z_{t-j}, \qquad j = 0, 1, 2, \tag{2.21}$$

3.0 IDENTIFICATION OF COST PERFORMANCE DATA

The initial step in the analyses of the cost performance time series (cumulative actual cost of work performed) is to determine if they are either stationary or non-stationary. For the two ongoing contracts used to illustrate the time series methodology, both series (ACWP) were plotted in an attempt to graphically detect any non-randomness or trend (nonstationarities). Figure 3.1 displays the cumulative ACWP for contracts T-22 (22-month duration) and N-37 (37-month duration). The graphs indicate that the T-22 and the N-37 data exhibit nonstationary behavior in their levels. In addition the N-37 data may exhibit nonstationary behavior in variability.

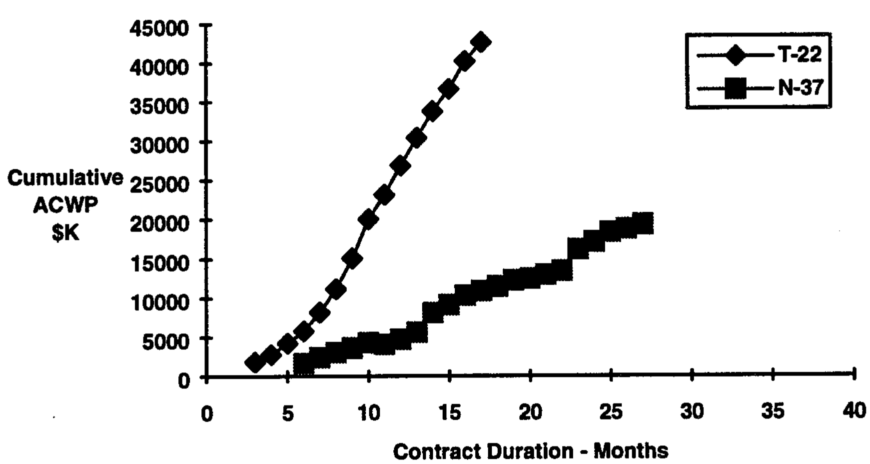

FIGURE 3.1 Cumulative Actual Cost of Work Performed for T-22 and N-37

The inference of the graphic displays was statistically tested for trend using *Kendall's Tau* test [3]. For further evidence, the sample autocorrelation functions were also calculated for the first- and second-order *difference filtered* data (see Figures 3.2 through 3.7).

Higher order filters are not indicated from Figure 3.1.

The critical value for *Kendall's Tau* test [4] at the $\alpha = .05$ level of significance is ± 1.645. The results of the tests are given below:

Table 3.1 Kendall's Tau Statistics for Trend, $Z_o = \pm 1.645$

Difference Filter Order	T-22 Data	N-37 Data	T-22 H_o:No Trend	N-37 H_o:No Trend
0	5.196	6.457	Reject	Reject
1	1.697	-.242	Reject	Accept
2	1.098	------	Accept	---------

This evidence indicates that both of the time series exhibit nonstationary properties and that first-differenced data for the N-37 series and second-differenced data for T-22 series showed no trend at the α=.05 level of significance. Also, the sample autocorrelation functions of both series failed to dampen rapidly (see Figures 3.2 and 3.3). This indicator further confirms that the T-22 and N-37 data are nonstationary (not in statistical equilibrium).

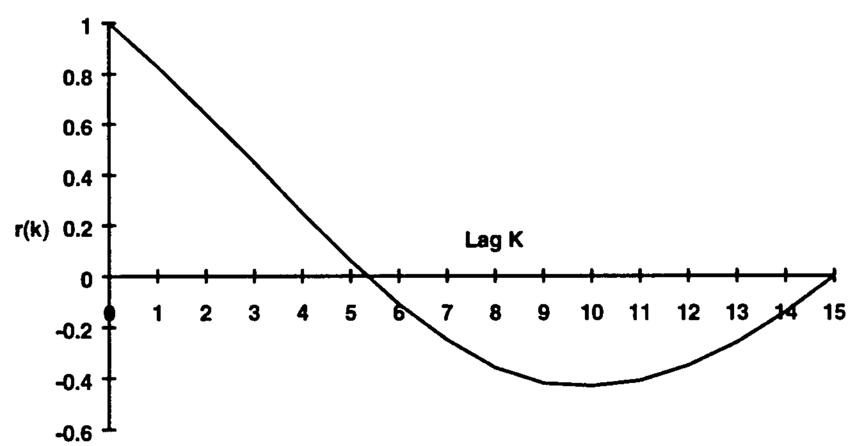

Figure 3.2 Sample Autocorrelation Function for T-22 Data

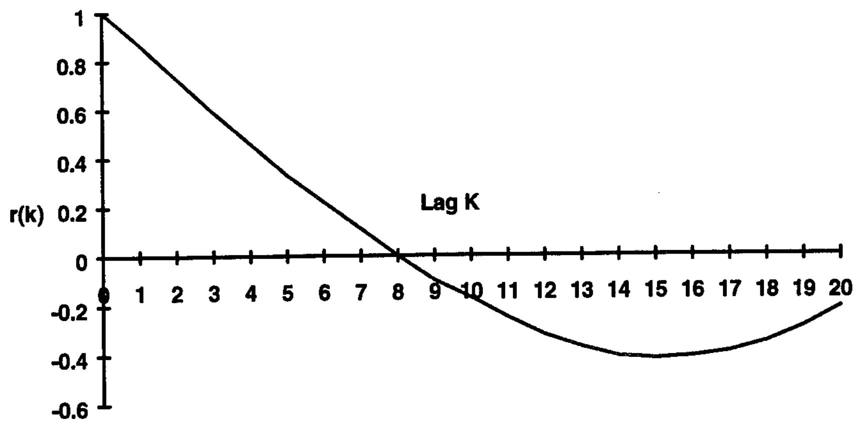

Figure 3.3 Sample Autocorrelation Function for T-37 Data

Figures 3.4 and 3.5 show the sample autocorrelation functions for the first-difference filtered data. Coincident with the results of the Kendall's Tau statistics, the first-filtered T-22 data do not dampen rapidly and the first-filtered N-37 data exhibit fairly rapid dampening. This warrants a look at the second-differenced information.

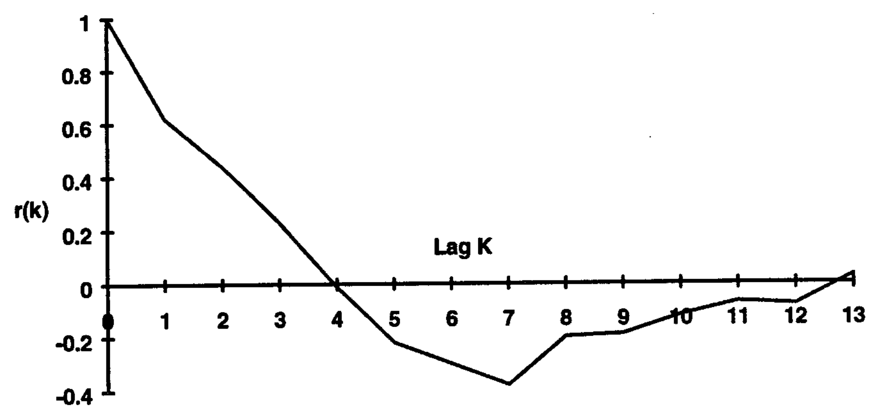

Figure 3.4 Sample Autocorrelation Function of the First-Difference T-22 Data

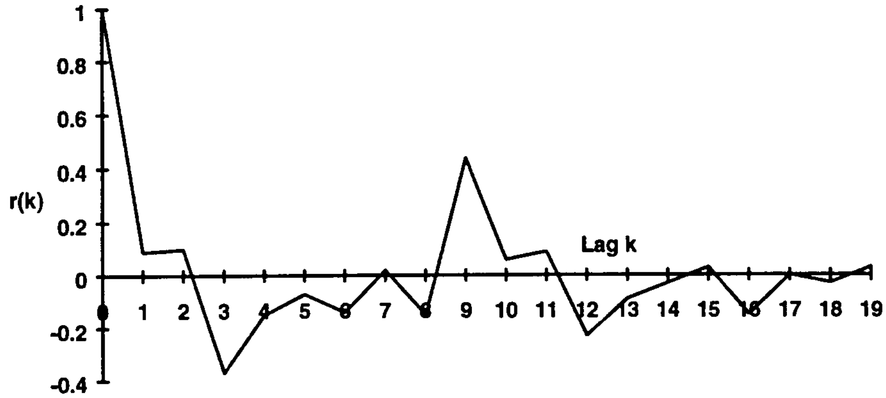

Figure 3.5 Sample Autocorrelation Function of the First-Difference T-37 Data

Graphic displays of the second-difference filtered data are shown in Figures 3.6 and 3.7. Here the T-22 data exhibit better dampening and the T-37 data do not.

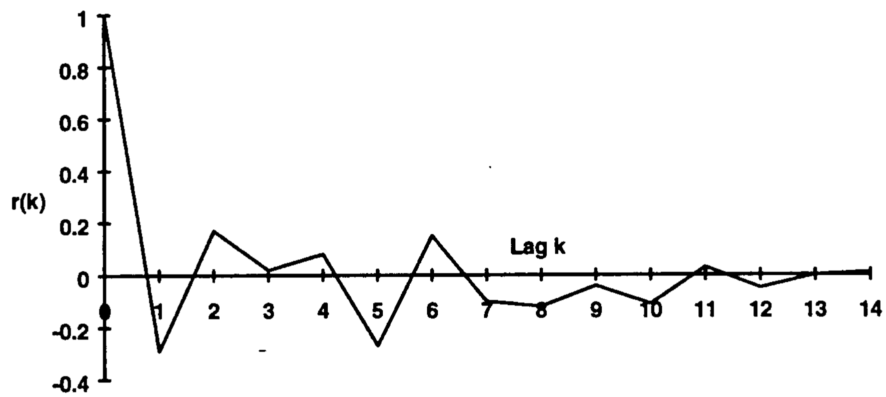

Figure 3.6 Sample Autocorrelation Function of the Second-Difference T-22 Data

Figure 3.7 Sample Autocorrelation Function of the Second-Difference N-37 Data

The characteristics of these displays and the results of the Kendall's Tau statistics indicate that the second-difference series for the T-22 data and the first-difference series for the N-37 data have reached statistical equilibrium. Therefore, with these filtered series, we can now proceed to fit forecasting models to the cumulative ACWP data.

3.1 Fitting the Models

To fit stationary stochastic models, either Autoregressive (AR(m)), Moving Average (MA(q)), or Autoregressive-Moving Average (ARMA(m,q)) to the filtered information as outlined in Section 2.0, estimates of the parameters $\alpha_1, \alpha_2, \cdots \alpha_m$, and $\beta_1, \beta_2, \cdots \beta_q$ for each process considered and for each order of the process must be made. It should be noted that, in practice, MA(q) models are useful for describing events that are effected by random events such as strikes and policy decisions [5]. In the case of the T-22 and N-37 data, moving average components would be indicators of the amount of risk associated with the work breakdown structure elements. Therefore, following the procedure outlined in Section 2.2, the parameters $\alpha_1, \alpha_2, \cdots \alpha_m$ and $\beta_1, \beta_2, \cdots \beta_q$ were estimated with the restriction that they lie between -1 and +1 to insure stationarity and/or invertibility of the filtered stochastic processes. The residual sums of squares were also computed and divided by the appropriate degrees of freedom to obtain the residual variance. Figures 3.8 and 3.9 show the residual variance as a function of model order (m,q) for the T-22 and N-37 data.

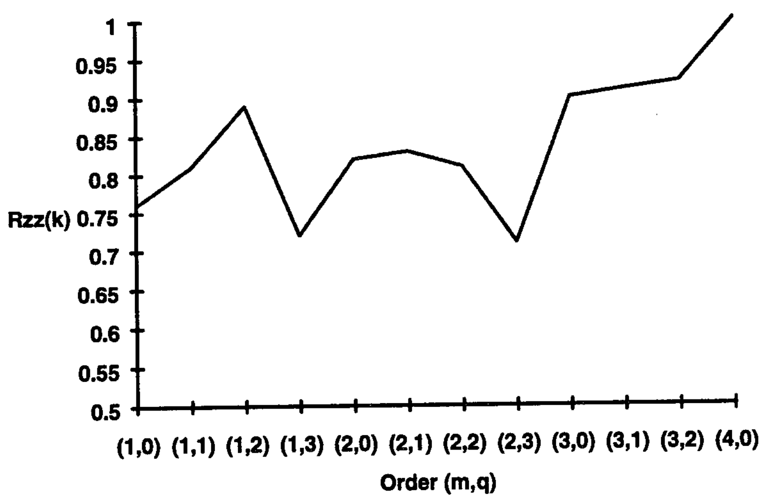

Figure 3.8 Model Order vs. Residual Variance for the T-22 Data

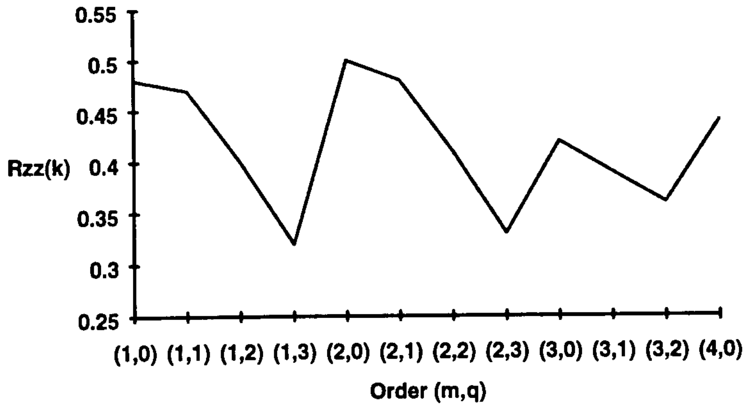

Figure 3.9 Model Order vs. Residual Variance for the N-37 Data

These displays indicate that the minimum residual variance criterion corresponds to an ARMA(2,3) model for the T-22 data, and to an ARMA (1,3) model for the N-37 data. The parameters for these models are shown in Table 3.2:

Table 3.2 Estimated Parameters for the T-22 and N-37 Models

	α_1	α_2	α_3	β_1	β_2	β_3
T-22	0.1875	0.3593		-0.8281	-0.3594	-0.1406
N-37	0.4375			-0.6719	-0.0312	-0.5625

Using the corresponding parameters the following difference equations were obtained for the filtered series:

For the T-22 data:

$$(w_t - 0.1145) = 0.1875(w_{t-1} - 0.1145) + 0.3594(w_{t-2}) + Z_t + 0.8281Z_{t-1} + 0.3594Z_{t-2} + 0.0.1406Z_{t-3} \quad (3.1)$$

and, for the N-37 data:

$$(y_t - 0.8379) = 0.4375(y_{t-1} - 0.8379) + Z_t + 0.6719Z_{t-1} + 0.0312Z_{t-2} + 0.5625Z_{t-3} \quad (3.2)$$

Since the T-22 data required a second-difference filter and the N-37 data required a first-order filter to be in statistical equilibrium, it is necessary to use *backwards* filters as outlined in Section 2. The filters are:

first order $\quad y_t = x_t - x_{t-1},$

second order $\quad w_t = x_t - 2x_{t-1} + x_{t-2}.$

These filters are inserted into equations (3.1) and (3.2) to obtain the appropriate forecasting models that will be used to characterize the T-22 and N-37 data. These equations become:

For the T-22 data:

$$\hat{x}_t = 2.188x_{t-1} - 1.016x_{t-2} - 0.531x_{t-3} + 0.359x_{t-4} + 0.052 + Z_t + 0.828Z_{t-1} + 0.359Z_{t-2} + 0.141Z_{t-3} \quad (3.3)$$

and, for the N-37 data:

$$\hat{x}_t = 1.438x_{t-1} - 0.438x_{t-2} + 0.471 + Z_t + 0.672Z_{t-1} + 0.031Z_{t-2} + 0.563Z_{t-3} \quad (3.4)$$

Setting the unknown Z_t's equal to their conditional expectations of *zero* and assuming the values $x_{t-1}, x_{t-2}, \cdots x_{t-m+d}$ have been realized, one can use equations (3.3) and (3.4) to simulate the observed data of both series. In addition, if t is replaced by t+ℓ in the above equations, one can then forecast ℓ steps (months) ahead, $\ell = 1, 2, \cdots L$, for both series. Figures 3.10 and 3.11 show the simulated data for the T-22 and N-37 data, which fits the data very well.

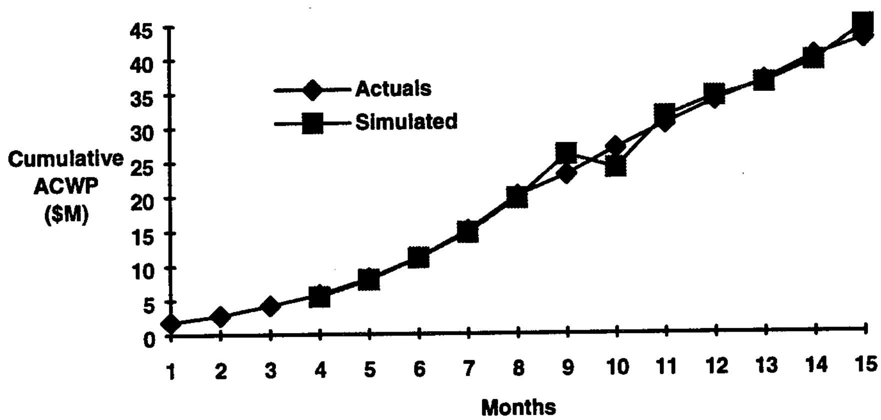

Figure 3.10 Simulated T-22 Series vs. the Actual Cumulative ACWP

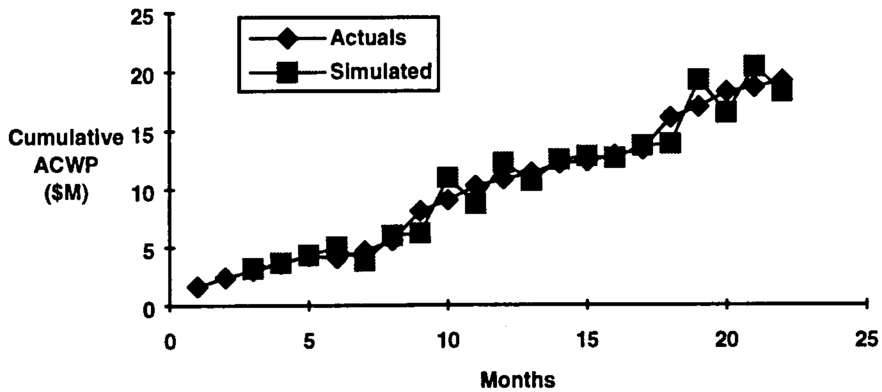

Figure 3.11 Simulated N-37 Series vs. the Actual Cumulative ACWP

Tables 3.3 and 3.4 show the ℓ step ahead forecasts at origin t=15 and t=22 for the T-22 and N-37 data, respectively. Errors, confidence bounds and updates of the forecasts are also shown.

Table 3.3 Forecasted Values of Cumulative ACWP for the T-22 Series at Origin x_{15} and Updates Under the Assumption That x_{16} Becomes Available.

Lead Time	Actuals ($M)	Forecast ($M)	95% Confidence Bounds ($M)	Error ($M)	Updated Forecast ($M)
Origin=x_{15}	*42.62*				
1	44.11	43.84	3.59	0.27	44.21
2	46.16	44.10	4.96	2.06	47.40
3	47.34	45.46	5.83	1.88	48.25
4	48.50	49.179	6.35	-0.68	48.44
5	50.00	53.95	6.82	-3.95	49.31
6	50.92	56.27	7.27	-5.35	49.84
7	51.63	52.71	7.68	-1.07	51.24

Table 3.4 Forecasted Values of Cumulative ACWP for the N-37 Series at Origin x_{22} and Updates Under the Assumption That x_{23} Becomes Available.

Lead Time	Actuals ($M)	Forecast ($M)	95% Confidence Bounds ($M)	Error ($M)	Updated Forecast ($M)
Origin=x_{22}	*19.25*				
1	19.78	15.92	5.22	3.86	18.87
2	20.02	18.96	5.84	1.06	19.63
3	20.15	19.75	5.84	0.40	19.75
4	20.55	23.04	-	-2.49	23.01
5	20.75	23.89	-	-3.14	23.85
6	21.00	22.77	-	-1.77	22.75
7	21.11	20.06	-	1.05	21.07
8	21.33	18.23	-	3.10	20.27
9	21.61	19.02	-	2.5	20.05
10	21.79	22.27	-	-0.48	22.26

Ordinarily, as ℓ increases, the forecasts become less accurate. However, the short term accuracy can be maintained by *updating* the forecasted values of the series as additional data become available. For example, the t=15 origin forecast of x_{17} of the T-22 data may be updated to become the t=16 origin forecast of x_{17} by adding a constant multiple of the one-step-ahead forecast error $\theta_\ell Z_{t+1} = \theta_1 Z_{16}$ to the t=15 origin forecast of x_{17}. The forecast error for this case is $Z_{16} = x_{16} - \hat{x}_{16}$, and, $\theta_\ell = \theta_1$, where $\theta_1 = \phi_1 - \hat{\beta}_1$ as explained in Section 2.0. The basis for updating the original forecasted data for ℓ steps ahead as additional data become available is:

$$\hat{x}_{(t+1)}(\ell) = \hat{x}_t(\ell+1) + \theta_\ell Z_{t+1} \tag{3.5}$$

4.0 SUMMARY AND CONCLUSIONS

The forecasting technique provided herein is intended to supplement the current forecasting capability in Performance Analyzer. Section 2.0 presented a procedural approach to time series modeling. Section 3.0 exercised the procedural approach developed to characterize the cumulative ACWP for two series of contractor performance data. Specifically, the cumulative ACWP was modeled for a short contract of twenty-two months duration (T-22) and one of thirty-seven months duration (N-37). Both series were shown to be nonstationary, and, following the procedural approach of Section 2.0, were characterized as ARMA(2,3) and ARMA(1,3), respectively (equations 3.3 and 3.4).

The results of these models compared to those of several Performance Analyzer models are shown in Table 4.1.

Table 4.1 Comparison of Time Series Results to Methods of Performance Analyzer (Estimates-at-Completion)

Contract	Origin t	3-Month Average ($M)	6-Month Average ($M)	Cum CPI ($M)	Time Series ($M)	Actual ($M)
T-22	16	58.16	55.26	54.22	52.71	51.63
N-37	22	23.27	22.28	20.95	22.27	21.79
	28	24.36	22.97	21.99	22.75	21.79
	31	23.65	23.28	22.25	22.27	21.79

In general, for the T-22 and N-37 series, the time series approach provides better forecasts than the methods of Performance Analyzer. It should be noted that the more observations that are available, the better the model. Analysts should be cautioned that this method is not recommended for contracts that are less than thirty months in duration. Though the T-22 contract is certainly less than thirty months, there is no guarantee that results of other contracts will be as good as this test case. It is also recommended that as many data points as possible be included in the modeling procedure. A good rule of thumb, in addition to the thirty-month recommendation, is to apply this technique to contracts at least 60% complete.

REFERENCES

1. Box, G.E.P., and G.M. Jenkins, Time Series Analysis, Forecasting, and Control. San Francisco, California: Holden-Day, 1970.

2. Jenkins, G.M., and D.G. Watts, Spectral Analysis and its Application. San Francisco, California: Holden-Day, 1968.

3. Kendall, M.G., and A. Stuart, The Advanced Theory of Statistics, Volume 3. London, England: Griffen, 1966.

4. Conover, W.J., Practical Nonparametric Statistics. New York: J. Wiley, 1971.

5. Levenbach, H., and J.P. Cleary, The Professional Forecaster. Belmont, California: Lifetime Learning Publications, 1982.